---

title: "Mini Project 1 (Solutions)"

author: "Buddhika Jayawardana"

date: "2/1/2020"

output: html\_document

---

### Section 1

1. In this problem we use the KNN classification method on the training data set and prodict the classes of the test data set.

a) knn() funtion was used for the K values = seq(1, 496, by = 1)

b) Training and test error rates were calculated for defferent K values and they were ploted against K values as follows.

```{r pressure, echo=FALSE}

```

c) Optimal value of K is found by finding the minimum value of test error rates and there were two K values that minimizes the test error rates, K = and K = . Associate training and test error rates were shown on the following table.

optimal.point

d) Since we had two K values that makes the test error rate minimize, corresponding decision boundaries were ploted seperatly as follows.

```{r pressure, echo=FALSE}

![alt text here](/prob1.partd.plot1.png)

![alt text here](\prob1.partd.plot2.png)

```

2. 1/5th of the data set e CIFAR-10 were used to classify using KNN method with K = 50, 100, 200, 300, 400.

a) Following table shows the error rates for defferent K values.

```{r pressure, echo=FALSE}

50 100 200 300 400

0.7030 0.7210 0.7355 0.7505 0.7565

```

b) Minimum error of test error rate was 0.7030 and the assosiate K value was 50. Bellow is the confusion metrix.

```{r pressure, echo=FALSE}

table(mod.test, y.test)

y.test

mod.test 0 1 2 3 4 5 6 7 8 9

0 112 16 34 9 6 12 6 12 18 11

1 0 0 0 0 0 0 0 0 0 0

2 12 11 45 32 17 33 22 33 5 6

3 0 0 0 4 0 0 0 1 1 0

4 37 64 108 109 121 105 109 97 28 71

5 0 0 0 1 0 13 0 1 1 0

6 5 35 14 29 12 19 38 14 7 12

7 0 0 1 1 0 3 2 10 0 5

8 49 65 13 17 12 9 9 32 130 101

9 0 7 1 2 0 2 1 4 2 14

```

c)

```{r pressure, echo=FALSE}

plot(ks, err.rate.test, xlab = "Number of nearest neighbors", ylab = "Test error rate", type = "b", ylim = range(c(err.rate.test)), col = "blue", pch = 20)

```